Forecasting the Widening Tuition Gap at the University of Virginia (1970–2030)

Samyu Krishnasamy(Leader), Hiba Khan, Alex Lozano
(Group 18 - Data Science Demons)

Motivation & Research Question

Context

UVA's tuition rates have consistently increased for both in-state and out-of-state students, but the gap between the two groups has expanded sharply since 2000.

Research Question

Can we accurately predict future tuition gaps between in-state and out-of-state students at UVA using historical data from 1970–2025, and does the model indicate a continued widening of this gap over time?

Hypothesis

If we achieve a MAPE >= 7% and an R^2 >= 0.8, we have successfully predicted the widening gap of In-State and Out-of-State Tuition.

Modeling Approach

Time-Series Trend Analysis

Confirmed the tuition gap is non-stationary and increasing over time

Holt's Linear Trend

Baseline model capturing steady upward growth

ARIMA

Modeled year-to-year momentum but struggled with large structural changes

Prophet (By Meta)

Captured shifts in growth rate (especially post-2000), leading to more stable forecasts

Data Acquisition and Overview

Data Source

- ★ Collected from the University of Virginia Office of Institutional Research & Analytics (IRA) public dashboards

Two original datasets:

- ★ In-State Tuition (1970–2024)
- ★ Out-of-State Tuition (1970–2024)

Merged into a single dataset:
Combined_UVA_Tuition_Cleaned.csv

Ethics

- ★ Publicly available, non-identifiable, and FERPA-compliant
- ★ No private student information was used

Data Format & Size

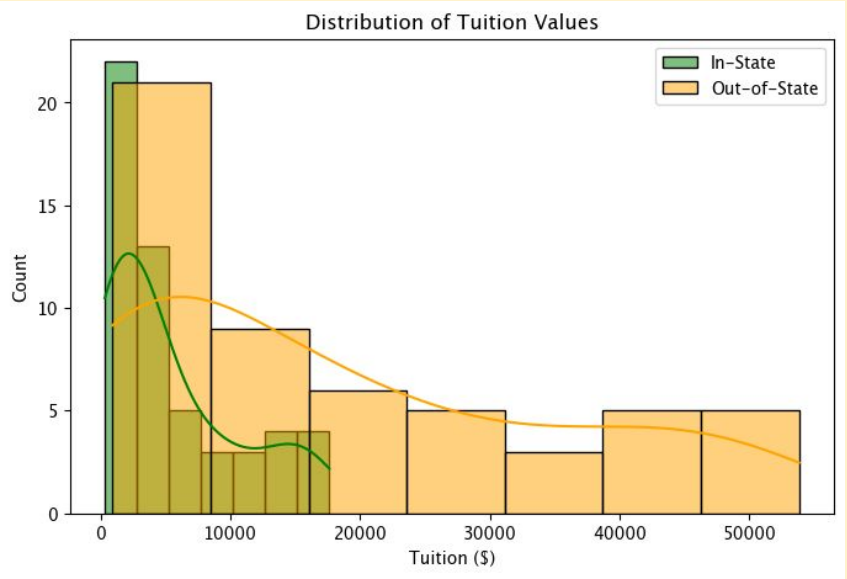
- ★ **Format**: CSV (tabular numeric data)
- ★ **Structure**: 55 yearly records (1970–2024), 12 columns
- ★ **Type**: Time-series (yearly observations)
- ★ **Software Used**: Python (pandas, matplotlib, statsmodels, Prophet)

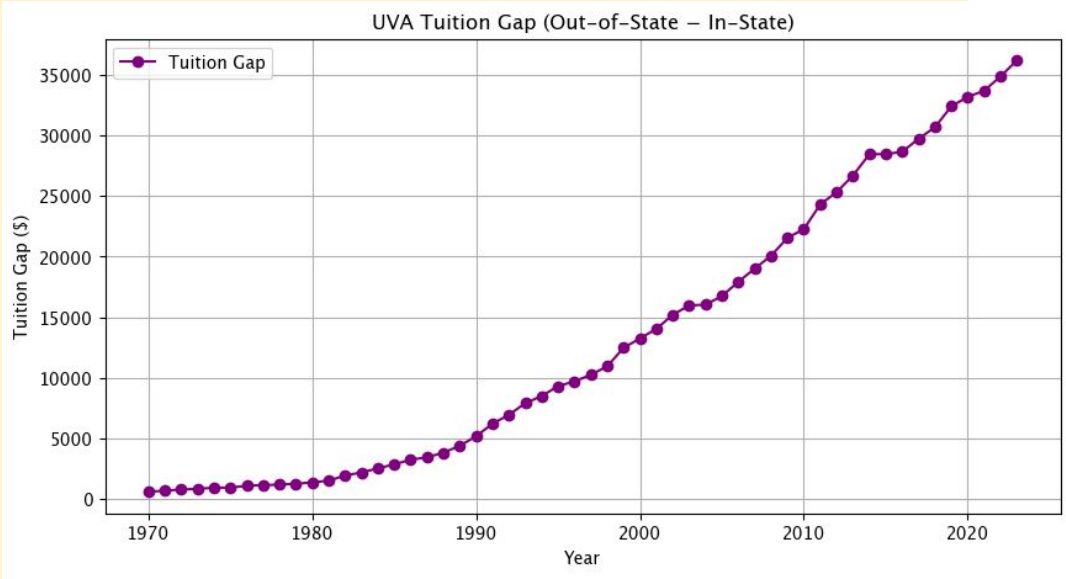
Key Variables

- ★ Year_Start: Academic year (time index)
- ★ Tuition_InState: Tuition for Virginia residents
- ★ Tuition_OutState: Tuition for non-residents
- ★ Tuition_Gap: Out-of-State − In-State tuition (modeling target)
- ★ Totals / Fees: Extra cost data, used for context only

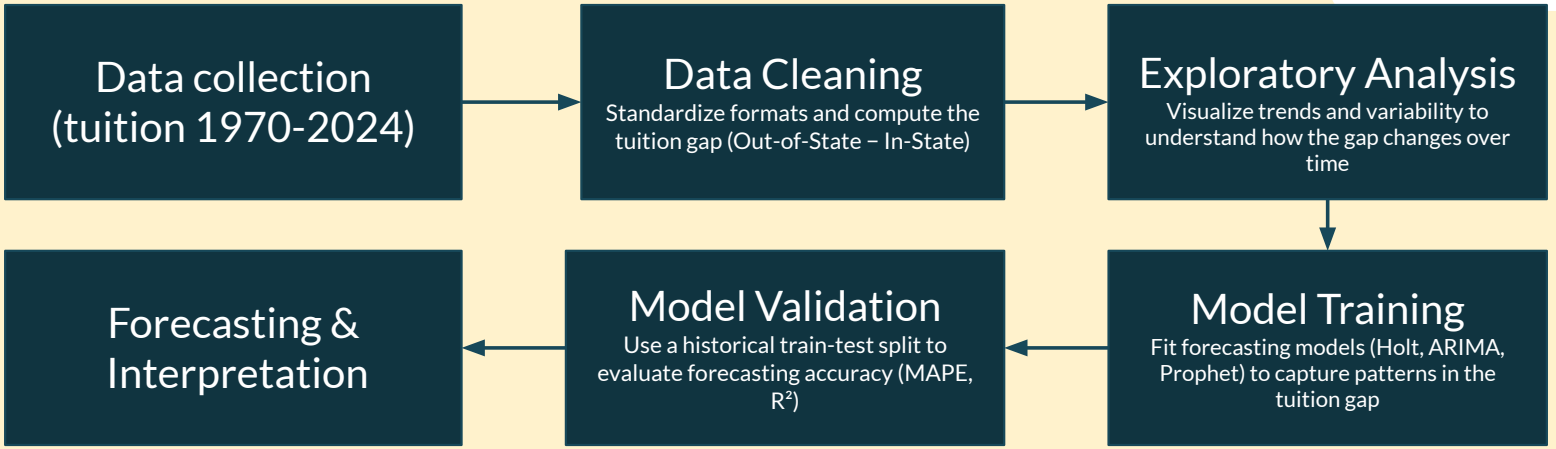
Variable Name	Description
Year_InState	Academic year label for in-state tuition data (e.g., "1970–71").
Residency_InState	Residency category for in-state students.
Totals_InState	Total annual cost for in-state students, including tuition and fees.
Required Fees_InState	Required annual fees for in-state students.
Tuition_InState	Annual tuition charged to in-state students.
Year_Start	Starting calendar year of the academic year (used for time-series indexing).
Year_OutState	Academic year label for out-of-state tuition data.
Residency_OutState	Residency category for out-of-state students.
Totals_OutState	Total annual cost for out-of-state students, including tuition and fees.
Required Fees_OutState	Required annual fees for out-of-state students.
Tuition_OutState	Annual tuition charged to out-of-state students.
Tuition_Gap	Difference between out-of-state and in-state tuition (OutState − InState).

Exploratory Charts





Analysis Plan and Justification



Tricky Analysis Decision

Limited GPA Data

- ★ GPA data only covers 2015–2024, too short for meaningful forecasting

Solution:

- ★ We decided to use tuition data from 1970–2024 instead, which gave us a much longer trend to study.

Changing Tuition Pattern

- ★ The tuition gap kept going up every year, not staying steady, meaning it wasn't stable over time

Solution:

- ★ Holt: Smoothed the line to see the overall trend.
- ★ ARIMA: Looked at the change from year to year.

Low Results

- ★ Holt & ARIMA: Had okay accuracy, but couldn't follow big shifts in the 2000s.
- ★ Noticed a trend shift in 2000's

Solution:

- ★ Researched models that could capture this
- ★ Prophet for accuracy + interpretability

Biases and Uncertainties

Data Bias

- ★ Older records may be less consistent due to changes in how tuition was reported.
- ★ Missing outside factors like inflation or funding can affect tuition trends.

Reducing Bias

- ★ Used official UVA Institutional Research data for accuracy.
- ★ Focused on tuition gap instead of raw tuition to reduce inflation effects

Uncertainties

- ★ Models perform well on historical data, but future tuition changes may vary because external factors (funding, enrollment, policy)

Results and Conclusions

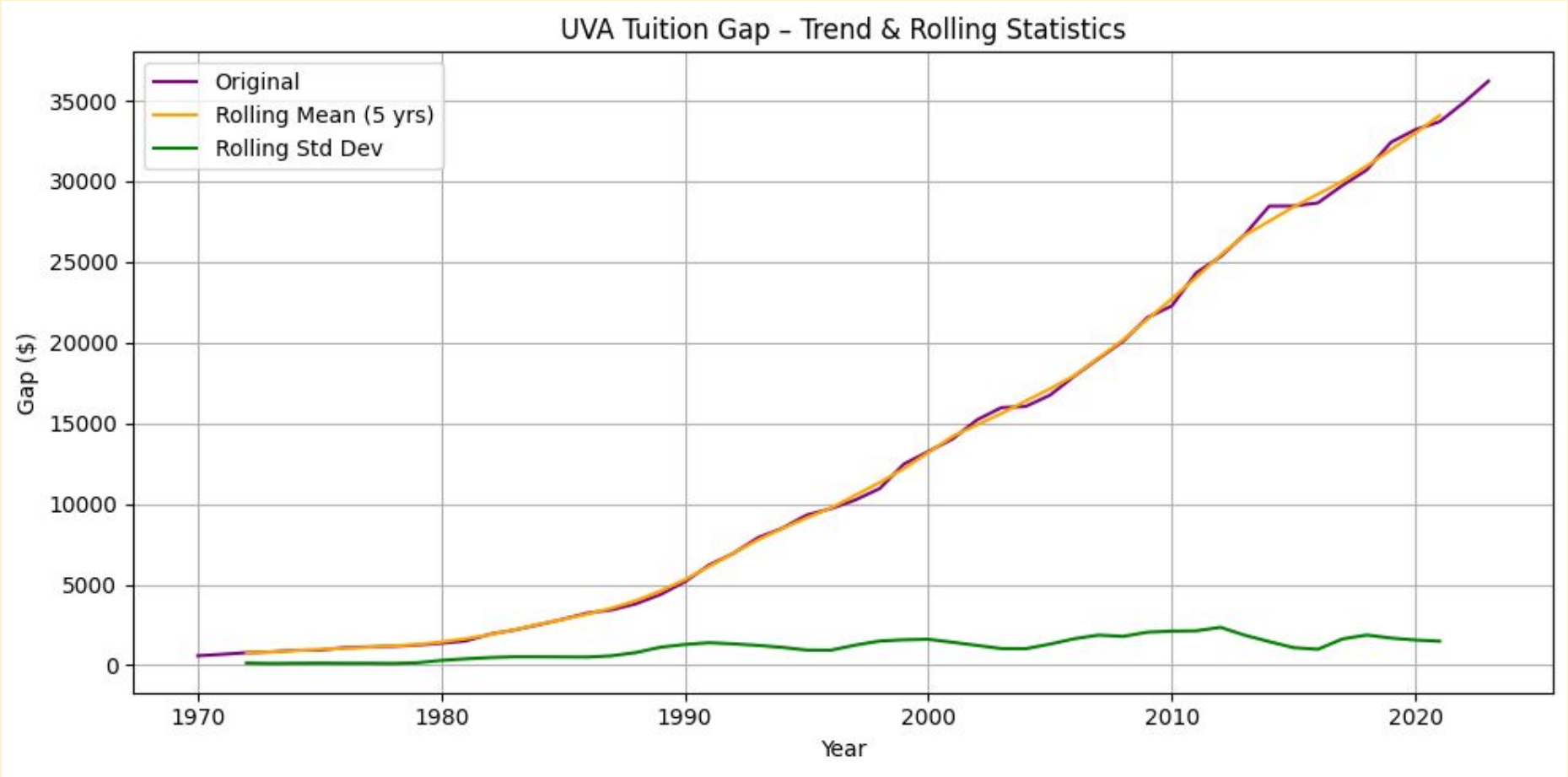
Goal: Forecast how UVA's tuition gap between in-state and out-of-state students will change through 2030.

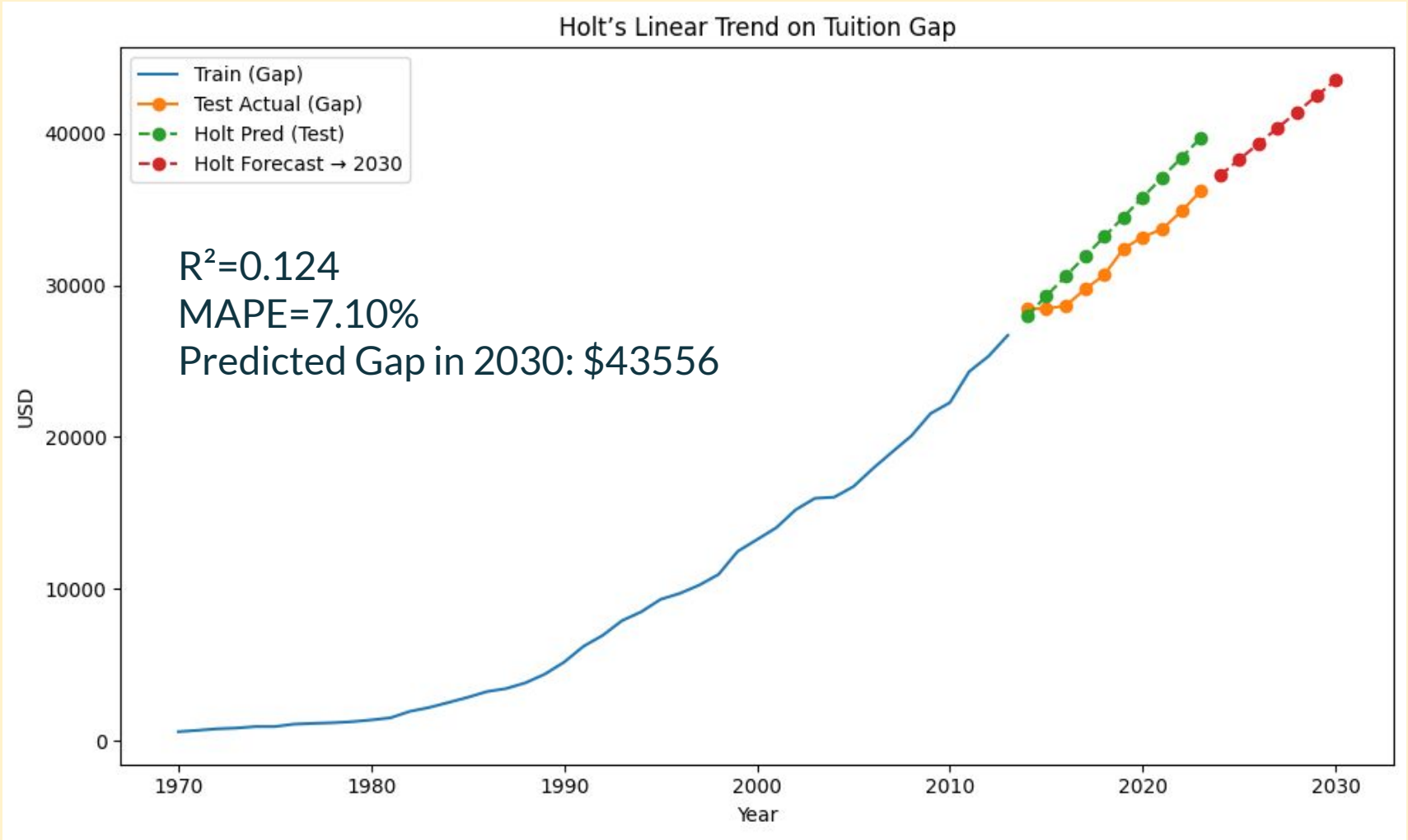
Key Findings

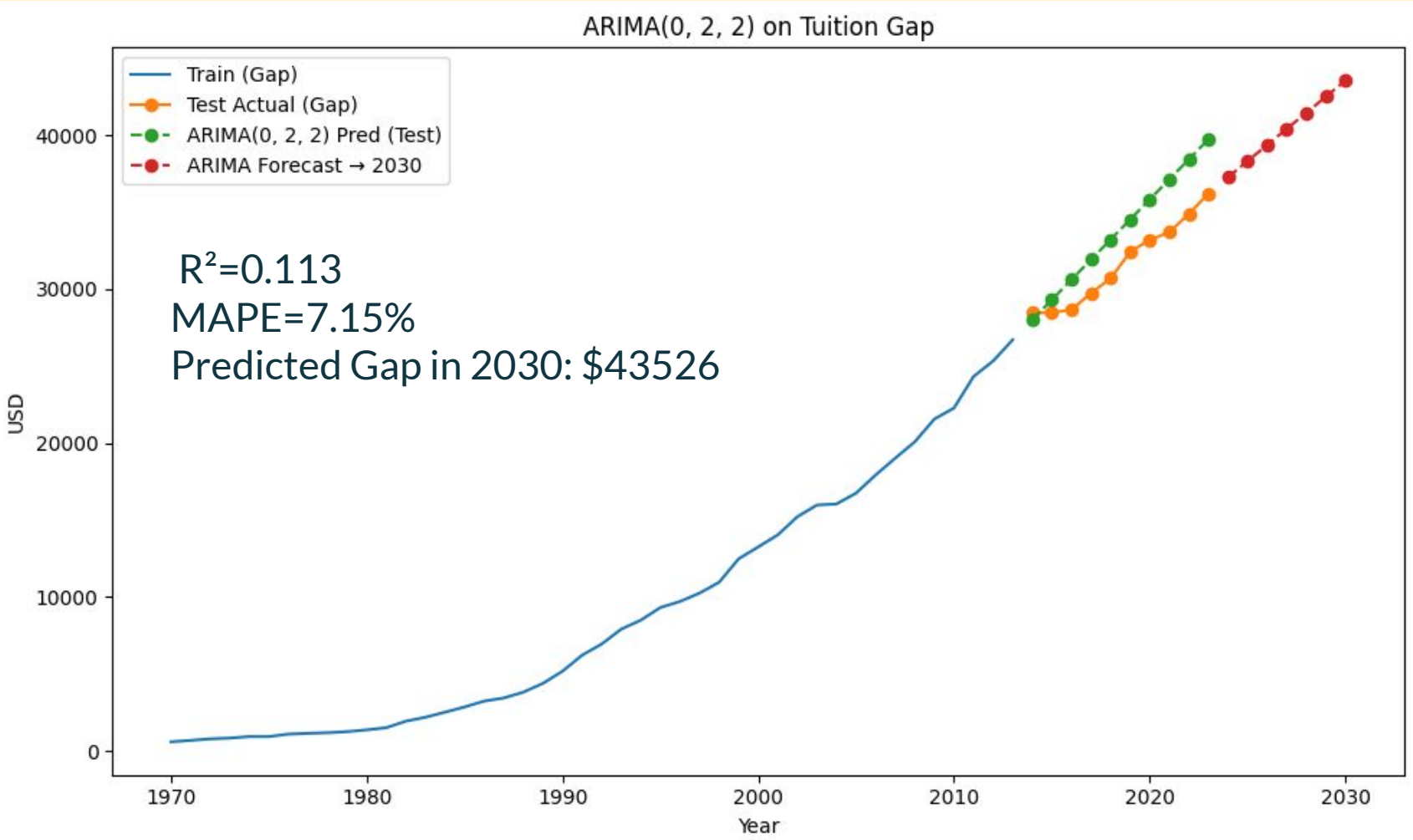
- ★ Tuition gap shows a steady upward trend since 1970 — the series is non-stationary.
- ★ Holt & ARIMA: captured short-term trends but missed major shifts in the 2000s.
- ★ Prophet: best performance — handled trend changes automatically.

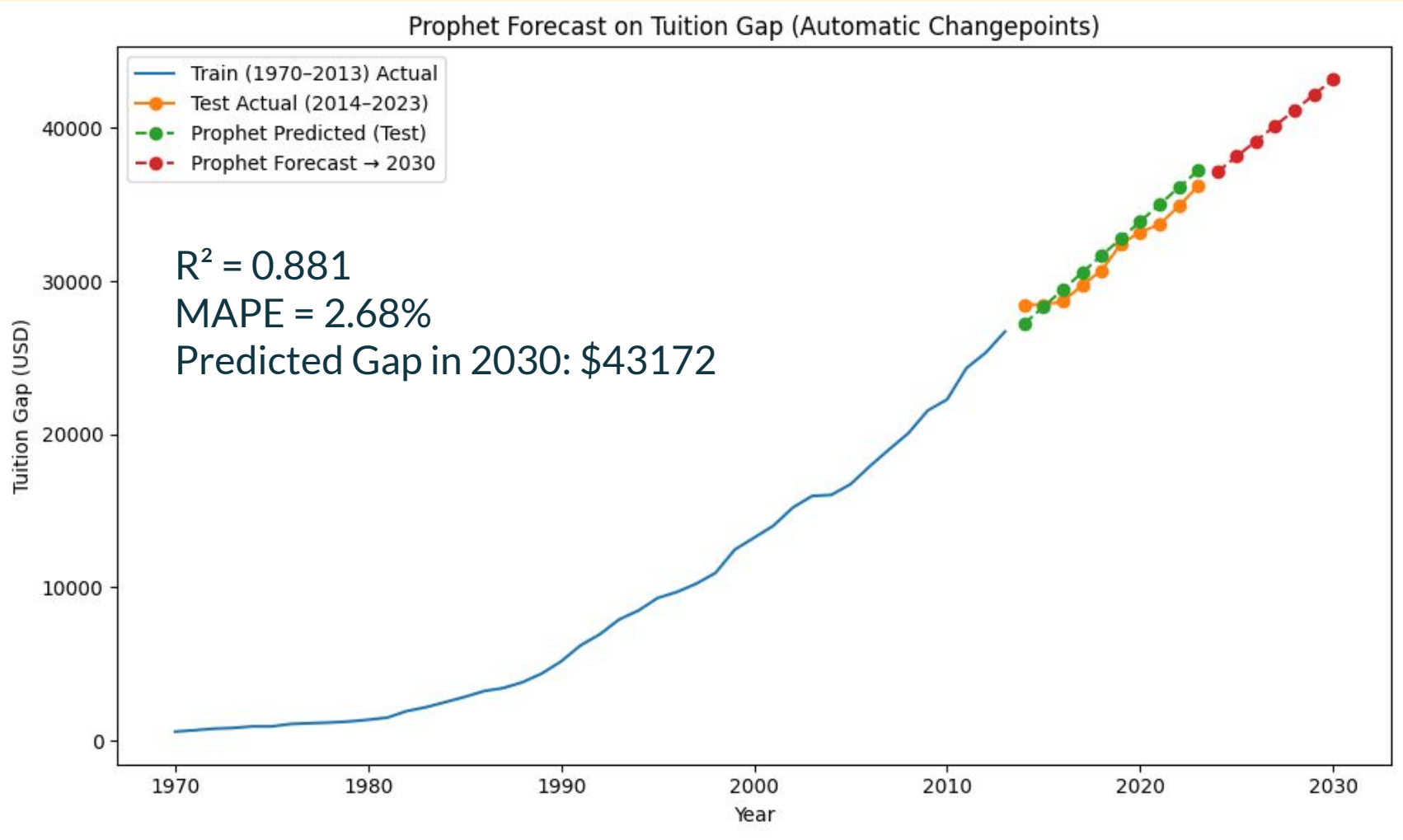
Model	R^2	MAPE	Direction (Last 5 yrs)
Holt Linear Trend	0.124	7.10 %	1.00
ARIMA (0, 2, 2)	0.113	7.15 %	1.00
Prophet	0.881	2.68 %	1.00

Conclusion: Hypothesis supported: models predict the tuition gap will continue widening through 2030. Prophet projects the gap to exceed $43,000 by 2030 if current trends persist.


UVA Tuition Gap – Trend & Rolling Statistics
Original
Rolling Mean (5 yrs)
Rolling Std Dev
Gap ($)
35000
30000
25000
20000
15000
10000
5000
0
1970
1980
1990
2000
2010
2020
Year


Holt’s Linear Trend on Tuition Gap
Train (Gap)
Test Actual (Gap)
Holt Pred (Test)
Holt Forecast → 2030
R²=0.124
MAPE=7.10%
Predicted Gap in 2030: $43556
USD
Year
0
10000
20000
30000
40000
1970
1980
1990
2000
2010
2020
2030


ARIMA(0, 2, 2) on Tuition Gap
Train (Gap)
Test Actual (Gap)
ARIMA(0, 2, 2) Pred (Test)
ARIMA Forecast → 2030
R²=0.113
MAPE=7.15%
Predicted Gap in 2030: $43526
USD
Year
0
10000
20000
30000
40000
1970
1980
1990
2000
2010
2020
2030


Prophet Forecast on Tuition Gap (Automatic Changepoints)
Train (1970–2013) Actual
Test Actual (2014–2023)
Prophet Predicted (Test)
Prophet Forecast → 2030
R² = 0.881
MAPE = 2.68%
Predicted Gap in 2030: $43172
Tuition Gap (USD)
0
10000
20000
30000
40000
1970
1980
1990
2000
2010
2020
2030
Year

Next Steps

Expand the Dataset

- ★ Incorporate additional UVA data such as average GPA, enrollment numbers, and state funding.
- ★ Combine academic and financial variables to study how performance and population changes relate to tuition trends.

Improve the Models

- ★ Adjust for inflation to track real cost growth.
- ★ Add these new variables as explanatory factors in models like ARIMAX or Prophet with regressors.

New Research Questions

- ★ Do changes in student performance or enrollment predict future tuition increases?
- ★ How strongly do policy or funding shifts influence UVA's tuition gap?

THANKS! (Questions?)

Resources:

GitHub:
https://github.com/samyukrishnasamy/DS-4002-Project2

University of Virginia Institutional Research and Analytics, Undergraduate GPA Dashboard, Charlottesville, VA: University of Virginia, 2024. [Online]. Available: https://ira.virginia.edu/. [Accessed: Oct. 17, 2025].

Open Data Commons, “Open Data Commons Attribution License (ODC-By 1.0),” Open Data Commons, 2024. [Online]. Available: https://opendatacommons.org/licenses/by/. [Accessed: Oct. 17, 2025].

S. J. Taylor and B. Letham, “Prophet: Forecasting at scale,” Facebook Research Blog, Feb. 2017. [Online]. Available: https://research.facebook.com/blog/2017/2/prophet-forecasting-at-scale/